

Maria Hauke-Kowalska, Wojciech Wesoly, Justyna Pac

Uniwersytet Przyrodniczy w Poznaniu, Katedra Hodowli Lasu, e-mail: maria.hauke@wp.pl

POZNANIE GENOMU KUBRYCKIEJ POPULACJI SOSNY ZWYCZAJNEJ

UNDERSTANDING OF THE GENOME OF SCOTS PINE INKUBRYK POPULATION

Słowa kluczowe: sekwencjonowanie nowej generacji, sosna, genom, NGS

Key words: New generation sequencing, pinus, genome, NGS

Abstract. In recent years there has been significant development of Next Generation Sequencing techniques that allow the sequencing of whole genomes. Department of Silviculture, University of Life Sciences in Poznan, in cooperation with the Laboratory of Genome Analysis, A. Mickiewicz University has an ongoing project of discovering the Scots pine genome of Kubryk population. This pine occurs in the Milicz Forest District, the area of Kubryk. In provenance studies this pine exceeds other populations in terms of the analyzed features.

WSTĘP

Koniec XX wieku przyniósł nam poznanie genomów kilkudziesięciu bakterii i organizmów wyższych. Na przełomie wieków ogłoszono sekwencję DNA człowieka. Fakt ten można uznać za krok milowy w dziejach ludzkości. Osiągnięcie to było efektem pracy wielkiego konsorcjum międzynarodowego - lista autorów publikacji liczy 130 nazwisk. W połowie lipca 2011 w *Nature* (tom 475, numer 7355 str. 189) opublikowana została praca: pt. „Sekwencja i analiza genomu ziemniaka produkującego bulwy” (Genome sequence and analysis of the tuber crop potato) autorstwa Konsorcjum Sekwencjonowania Genomu Ziemniaka (Potato Genome Sequencing Consortium — PGSC). W skład tego międzynarodowego Konsorcjum weszły 32 Zespoły z 14 krajów (w tym z Polski) będących znaczącymi producentami i uczestnikami programów hodowli ziemniaków [Gromadka i in. 2011]. W tym samym czasie ukazała się publikacja o genomie ogórka, w której lista autorów zawiera tylko 20 nazwisk z trzech jednostek naukowych, z tego 16 Polaków z jednej katedry [Wóycicki i in. 2011]. Liczba poznanych genomów roślinnych lawinowo rośnie. Wśród nich są przede wszystkim te, które służą jako organizmy modelowe lub są ważne w gospodarce człowieka [Malepszy i in. 2012]. Obecnie znane są genomy roślin tj. rzodkiewnika, kukurydzy ogórka, ziemniaka, pomidora, ryżu i innych, natomiast z drzew leśnych opisano genom topoli (*Populus trichocarpa*). W opracowaniu są genomy innych gatunków np. *Pinus tadea*, *Picea abies* [Nystedt i in., 2013] i innych roślin drzewiastych (baza danych Dendrome Project).

Dzięki rozwojowi nowoczesnych technik analizy genomu, w tym sekwencjonowania nowej generacji, jest możliwość analizy całych genomów przy znacznie mniejszym zaangażowaniu kosztów i zespołów badawczych. W Katedrze Hodowli Lasu Uniwersytetu Przyrodniczego w Poznaniu, przy współpracy z Środowiskowym Laboratorium Analizy Genomu Uniwersytetu im. A. Mickiewicza w Poznaniu, jest realizowany projekt sekwencjonowania genomu sosny zwyczajnej populacji kubryckiej.

SOSNA ZWYCZAJNA

Gatunek *Pinus sylvestris* L. ma największy zasięg spośród innych gatunków z rodzaju *Pinus* w Europie i Azji. Optymalne warunki wzrostu sosna znajduje w Polsce i występuje na całym obszarze kraju z wyjątkiem południowo-wschodniego krańca Bieszczad. Pospolicie na niżu, rzadziej na pogórzu, najrzadziej w górach. Udział jako panującego gatunku w polskich lasach obecnie wynosi około 70% [Matuszkiewicz i in. 2012]. Nie odznacza się najwyższą produktywnością i ustępuje pod względem przyrostu drzewostanom innych gatunków. Powszechne występowanie lokuje drzewostany sosnowe na pierwszej pozycji pod względem znaczenia gospodarczego [Białobok i in. 1993]. Do dużego udziału gatunków iglastych przyczyniło się również ich preferowanie, począwszy od XIX w., przez przemysł drzewny.

Sosna znalazła w Polsce najkorzystniejsze warunki klimatyczne oraz siedliskowe w swoim zasięgu, dzięki czemu zdołała wytworzyć wiele cennych ekotypów. Do najbardziej cenionych należą populacje z północno-wschodniej oraz południowo-zachodniej części kraju. Możemy tutaj przytoczyć „sosnę mazurską” ekotypy z Pisz, Guzianki i Rucianego oraz ekotypy pochodzące z Bolewic i Rychtała. Długoletnie badania proveniencyjne sosny zwyczajnej w Polsce dostarczyły szereg informacji na temat przystosowania i wartości produkcyjnych rodzimych populacji tego gatunku. Trudnością napotykaną w badaniach proveniencyjnych drzew jest ich długowieczność oraz tzw. zmienność ciągła gatunku, czyli występowanie form przejściowych na niewielkich obszarach [Matras 1996].

Jednym z najbardziej interesujących ekotypów sosny zwyczajnej występującej na terenie Nadleśnictwa Milicz, w obrębie Kubryk, jest tzw. sosna kubrycka - posiadająca status zagrożona, do ochrony biernej.

Obręb Kubryk leży w II Krainie Wielkopolsko- Pomorskiej w 9 Dzielnicy Kotliny Żmigrodzko-Grabowskie, jak również w V Krainie Śląskiej w 2 Dzielnicy Wrocławskiej w mezoregionie wzgórz Trzebnicko-Ostrzeszowskich.

Przed II wojną światową obręb Kubryk Nadleśnictwa Milicz należał do niemieckich lasów państwowych – Nadleśnictwo Kuhbruck. Odnośnie prowadzenia gospodarki leśnej do 1945r. brak jest sprawdzonych danych. Wiadomo jedynie, że użytkowanie drzewostanów prowadzono przeważnie zrębami

zupelnymi na wiekszych powierzchniach odnawianych sosna. Sosna z niemieckiego Nadleśnictwa Kubryk była bardzo ceniona przez leśników niemieckich. W większości sosna ta została wycięta po koniec II wojny światowej. Obecnie jest brak jakichkolwiek danych o pochodzeniu tej sosny. Możemy założyć dwie hipotezy dotyczące jej pochodzenia:

- Jest to ekotyp, który dostosował się do warunków na drodze ewolucji.
- Jest to ekotyp wprowadzony w ramach tzw. mieszanki pruskiej.

Pula genetyczna tej sosny jest bardzo zawężona. **Ekotyp kubrycki wyróżnia się wyjątkową długowiecznością, strzelistą i smukłą formą pni oraz wysokiej jakości drewnem. Ponadto w doświadczeniu proweniencyjnym opisanym przez Barzdajna [2006] za najlepszą populację pod względem analizowanych cech pierśnic (skorygowanych liczbą drzew na poletkach), średnich wysokości, sumy powierzchni przekrojów pierśnicowych oraz miąższości strzał, uznano populację Milicz tzw. sosnę kubrycką.**

SEKWENCJONOWANIE GENOMÓW

Informacja genetyczna jest niezbędnym składnikiem każdego organizmu żywego. Dzięki zapisowi informacji genetycznej organizm żyje, rozwija się oraz jest podatny lub odporny na szereg czynników środowiska. Całość informacji genetycznej zawartej w każdej komórce żywej organizmu nazywa się genomem. Termin ten po raz pierwszy został zaproponowany w 1920 roku przez niemieckiego botanika Hansa Winklera i oznacza wszystkie kodujące i niekodujące sekwencje DNA zawarte w haploidalnej liczbie chromosomów.

Sekwencjonowanie polega na określeniu sekwencji zasad w DNA. Historia rozwoju metod sekwencjonowania jest dość krótka. W 1953 roku Watson i Crick opracowali model podwójnej helisy DNA. W 1977 roku Allan Maxam i Walter Gilbert opracowali metodę sekwencjonowania DNA przez chemiczną degradację. W tym samym roku Fred Sanger zaprezentował metodę sekwencjonowania DNA metodą syntezy katalitycznej. Trzy lata później Fred Sanger i Walter Gilbert otrzymali Nagrodę Nobla za swoje badania [Barciszewski, Markiewicz 2002]. Metoda dideoksy Sangera jest najbardziej klasyczną i najszerzej stosowaną metodą sekwencjonowania. Obecnie procedura ta została odpowiednio zmodyfikowana, aby przyspieszyć reakcję i polepszyć jakość wyników. Ponadto została zautomatyzowana. Sekwencja odczytywana jest za pomocą laserowych detektorów, a jej zapis jest bezpośrednio wpisywany do pamięci komputera, eliminując tym samym błędy wynikające z niedokładnego odczytu wyniku przez człowieka. Umożliwia ona odczyt sekwencji nukleotydów od około 100 do 1000 par zasad. Z tego powodu idealnie nadaje się do analiz produktów otrzymanych dzięki PCR (łańcuchowej reakcji polimerazy). Przybliżony czas odczytu 1000 par zasad to ok. 125 min. Ta zautomatyzowana technika pozwoliła na sekwencjonowanie całych genomów organizmów żywych, takich jak wirusy (np. HIV) czy bakterii (np. *Helicobacter pylori*).

Wysoki koszt tradycyjnego sekwencjonowania wymusił rozwój technologii sekwencjonowania nowej generacji o wysokiej przepustowości. Od 2005 roku dostępne są na rynku techniki sekwencjonowania nowej generacji (ang. Next Generation Sequencing- NGS) pozwalające na równoległe prowadzenie reakcji sekwencjonowania i generowanie tysięcy lub miliony sekwencji jednocześnie. Technologie sekwencjonowania nowej generacji o wysokiej przepustowości mają na celu przede wszystkim znaczną redukcję kosztów sekwencjonowania DNA, ale także przyspieszenie procesu masowego sekwencjonowania. Stosując tradycyjne metody, ten cel nie jest możliwy do osiągnięcia.

Jedną z najbardziej ekonomicznych technologii wykorzystywanych w NGS jest technologia Solexa-Illumina. Opracowała ona metodę sekwencjonowania opartą na terminatorach z odwracalnie usuwalnymi barwnikami fluorescencyjnymi. Proces wygląda dość skomplikowanie, jednak jest dość prosty. Na odpowiednio przygotowanej wcześniej płytce (flow cell) umieszcza się jednoniciowe fragmenty DNA, które następnie zaginane są w mostek. Mostek ten przyłącza się do powierzchni dzięki adapterom dołączonym do fragmentów DNA w fazie przygotowania bibliotek, a następnie powielane są za pomocą polimerazy, aby uzyskać 20-30 kopii tego samego insertu (tzw. amplifikacja mostkowa, ang. bridge amplification). Tylko jeden typ nukleotydu może być przyłączony w danym cyklu. Pozostałe niezwiązane nukleotydy są wymywane. Kamera cyfrowa rejestruje powstały obraz fluorescencyjny znakowanych nukleotydów. Następnie barwniki fluorescencyjne zostają chemicznie usunięte z DNA, umożliwiając zajęcie kolejnego cyklu i rejestrację kolejnego obrazu. Zbiór wszystkich fotografii pozwala odtworzyć badane sekwencje DNA w sekwenatorze [Mardis 2008; www.illumina.com].

Zanim zaczniemy sekwencjonowanie koniecznym jest przygotowanie bibliotek z DNA genomowego lub RNA, które przepisujemy na cDNA. Podczas przygotowania bibliotek możemy wyróżnić następujące etapy: (1) – fragmentacja lub/i dobór fragmentów o docelowej długości, (2) – dołączenie adapterów oligonukleotydowych do fragmentów DNA, (3) – oznaczenie ilościowe produktu końcowego biblioteki do sekwencjonowania [Head i in., 2013]. Kolejnym krokiem jest sekwencjonowanie, które jest pierwszym etapem na drodze do odczytania kolejności nukleotydów w sekwencji DNA. Od wyników sekwencjonowania zależy dalsza praca nad odczytaniem kodu genetycznego. Dzięki sekwencjonowaniu są odczytywane sekwencje DNA długości zazwyczaj do kilkuset par nukleotydów. Po tym etapie uzyskujemy olbrzymią liczbę danych. Odczytywane sekwencje są jednak bezwartościowe bez odpowiednich narzędzi informatycznych do ich analizy. Obecnie etapem ograniczającym uzyskiwanie znaczących informacji staje się nie samo sekwencjonowanie genomów, lecz ich analiza [McPherson 2009].

Kolejnym, najbardziej czasochłonnym w całej procedurze krokiem jest złożenie zyskanych odczytów w sekwencję DNA, a później rozszyfrowanie „treści” zakodowanej w odczytanej sekwencji i wyciąganie wniosków wyjaśniających

zjawiska biologiczne. „Składanie” krótkich sekwencji w całość odbywa się za pomocą dwóch zasadniczych typów algorytmów: *assembly* i *alignment* [Flicek i Birney 2009]. Składanie sekwencji *de novo* (assembly) opiera się na identyfikacji kolejnych częściowo nakładających się fragmentów. Należy podkreślić, że nowo zsekwencjonowany genom staje się genomem referencyjnym, nieodzwierciedlającym ani wewnątrzgatunkowej zmienności, ani stanu heterozygotyczności wewnątrz zsekwencjonowanego genomu [Malepszy i in. 2012]. Podejście *alignment* ma zastosowanie wtedy, gdy dysponujemy sekwencją genomu innego organizmu tego samego gatunku lub też gatunku spokrewnionego. Pojedyncze odczyty są wówczas przypisywane do odpowiednich regionów genomu referencyjnego. Dostęp do genomu referencyjnego radykalnie ułatwia i obniża koszty sekwencjonowania kolejnych przedstawicieli gatunku (tzw. resekwencjonowanie), umożliwiając zarówno badanie polimorfizmu, jak też użycie sekwencjonowania jako metody diagnostycznej.

Porównywanie sekwencji nukleotydowych jest obecnie w biologii jednym z najbardziej rozpowszechnionych i podstawowych podejść badawczych. Porównywanie sekwencji z wielu genomów ułatwia poznanie genów i określenie ich struktury. Dzięki analizie porównawczej możemy wskazać interesujące nas geny np. odpowiedzialne za przyspieszony wzrost. Szczegółowe analizy kierunkowe genotypów pozwalają na odpowiedź, czy dany osobnik posiada określoną cechę determinowaną przez analizowany gen. Obecnie, w celu analizy np. zmienności genetycznej albo odporności na dany czynnik, przeprowadza się dwie odrębne analizy. W przypadku zastosowania nietrafionego markera przeprowadza się następne badanie, co wiąże się z kolejnymi dużymi kosztami. Jednorazowe zsekwencjonowanie genomu pozwala na odczyt wszystkich możliwych sekwencji determinujących interesujące leśnika cechy. Umożliwia także wielokrotne wykorzystanie danych do analiz genetycznych w późniejszym czasie. Takie podejście jest bardzo dobrą podstawą do rozwoju selekcji leśnej na poziomie molekularnym, w przyszłości. Znajomość pełnego genomu organizmu żywego umożliwia zrozumienie wielu molekularnych mechanizmów dotyczących ich funkcjonowania i ewolucji. Ponadto znajomość całej cząsteczki DNA umożliwia badanie zmienności genetycznej oraz pozwala na szukanie mutacji powstałych w genomie, które wpływają na ową zmienność. Wielkość zmienności daje nam obraz pokrewieństwa pomiędzy organizmami i pozwala niemalże „ogłądać ewolucję” [Kasprzak, Świerszcz 2009].

PODSUMOWANIE

Nowoczesne leśnictwo oparte jest na doświadczeniach hodowlanych oraz selekcji gatunków drzew leśnych. Trudnością napotykaną w badaniach proweniencyjnych drzew jest ich długowieczność. W tym kontekście coraz większego znaczenia nabiera możliwość wykonania analiz genetycznych

pozwalających na skrócenie czasu otrzymania wyników badań. Jednym z najbardziej rozpowszechnionych badań są analizy markerów DNA. Markery opisują zmiany w sekwencji nukleotydów DNA jądrowego, mitochondrialnego i chloroplastowego.

Zsekwencjonowanie genomu sosny zwyczajnej populacji kubryckiej oraz populacji rychtalskiej będącej krajowym standardem w testowaniu potomstwa pozwoli nam na uzyskanie genomu referencyjnego, który będzie wykorzystywany w dalszych analizach. Poznanie sekwencji DNA da odpowiedź dotyczącą pochodzenia sosny kubryckiej. Ponadto, w kolejnych etapach będzie podjęta próba wytypowania sekwencji genomowych odpowiedzialnych za przyspieszony wzrost.

Poznanie sekwencji genomu sosny zwyczajnej, czyli „planu działania” wskazującego na mechanizmy rządzące całym organizmem, pozwoli genetykom na wyselekcjonowanie genotypów o pożądanych przez leśników cechach.

LITERATURA

- Barciszewski J., Markiewicz W. T. (2002). *Kwasy nukleinowe. Kod genetyczny*. W: Koroniak H., Barciszewski J. (red.). Na pograniczu chemii i biologii. T. V. PWN.
- Barzdajn W., (2006). *Zmienność cech taksacyjnych sosny zwyczajnej (Pinus sylvestris L.) polskich pochodzeń w doświadczeniu proweniencyjnym z 1985 roku w Nadleśnictwie Zielonka*. Sylwan 150 (1): 8–19.
- Biabok S., Boratyński A., Bugała W. (1993). *Biologia Sosny zwyczajnej*. Polska Akademia Nauk-Institut Dendrologii w Kórniku. Polska.
- Bogunic F., Muratovic E., Brown S.C., Siljak-Yakovlev S., (2003): *Genome size and base composition of five Pinus species from the Balkan region*. Plant Cell Rep. 22:59-63
- Flicek P., Birney E., (2009). *Sense from sequence reads: methods for alignment and assembly*. Nat. Methods., 6(11 Suppl), S6-S12.
- Gromadka R., Gawor J., Szczęśny P., Zagórski W. (2011). *Kolejny wielki genom poznany przy udziale polskich laboratoriów. Genom ziemniaka zsekwencjonowany*. Kosmos 60, 491–497.
- Head S. R., Komori H. K., LaMere S. A., Whisenant T., Van Nieuwerburgh F., Salomon D. R., Ordoukhanian, P. (2013). *Library construction for next-generation sequencing: overviews and challenges*. BioTechniques 56(2), 61-4.
- Illumina www.illumina.com/documents/products/techspotlights (dostęp 07.12.2014)
- Kasprzak M., Świercz A. (2009). *Sekwencjonowanie i asemblacja DNA –podejścia, modele grafowe i algorytmy*. Kosmos 58 (1-2): 17-28
- Malepszy S., Przybecki Z., Kowalczyk C. 2012. *Sekwencjonowanie genomów staje się nowym składnikiem postępu w hodowli roślin*. Kosmos. 61(3): 467-475.
- Mardis R. E., (2008). *Next-generation DNA sequencing methods*. Annual Review of Genomics and Human Genetics, 9:387–407.
- Matuszkiweicz W., Sikorski P., Szwed W., Wierzba M. (2012). *Zbiorowiska Roślinne Polski. Lasy i zarośla*. Wydawnictwo Naukowe PWN Warszawa 2012, ss. 27-29
- Maxam A. M., Gilbert W. (1977). *A new method for sequencing DNA*. Proceedings of the National Academy of Sciences of the United States of America, 74(2), 560–564.
- McPherson J. D., (2009), *Next-generation gap*. Nat. Methods., 6(11 Suppl), 2-5

- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., ... & Thompson, S. L. (2013). *The Norway spruce genome sequence and conifer genome evolution*. *Nature* 497(7451), 579-584.
- The Potato Genome Sequencing Consortium (2011). *Genome sequence and analysis of the tuber crop potato*. *Nature* 475, 189–195.
- Wakamiya I., Newton R.J., Johnston S.J., Price J.H., (1993): *Genome size and environmental factors in the genus Pinus*. *American Journal of Botany* 80:1235–1241
- Wóycicki, R., Witkowicz, J., Gawroński, P., Dąbrowska, J., Lomsadze, A., Pawełkowicz, M., Siedlecka, E., Yagi, K., Pląder, W., Seroczyńska, A., Śmiech, M., Gutman, W., Niemirowicz-Szczytt, K., Bartoszewski, G., Tagashira, N., Hoshi, Y., Borodovsky, M., Karpiński, S., Malepszy, S., Przybecki, Z. (2011). *The Genome Sequence of the North-European Cucumber (Cucumis sativus L.) Unravels Evolutionary Adaptation Mechanisms in Plants*. *PLoS ONE* 6(7): e22728.

STRESZCZENIE

W ostatnich latach coraz częściej dowiadujemy się o poznaniu genomu kolejnego organizmu. Od 10 lat dostępne są techniki sekwencjonowania nowej generacji pozwalające na równoległe prowadzenie reakcji. Wysokopręciowe technologie mają na celu przede wszystkim znaczną redukcję kosztów, a także przyspieszenie procesu masowego sekwencjonowania.

Dzięki rozwojowi nowoczesnych technik sekwencjonowania jest możliwość analizy całych genomów, przy relatywnie niewysokich kosztach. W Katedrze Hodowli Lasu Uniwersytetu Przyrodniczego w Poznaniu, przy współpracy z Środowiskowym Laboratorium Analizy Genomu Uniwersytetu im. A. Mickiewicza w Poznaniu, jest realizowany projekt poznania genomu sosny zwyczajnej populacji kubryckiej. Sosna ta występuje na terenie Nadleśnictwa Milicz, w obrębie Kubryk. W badaniach proveniencyjnych przewyższa inne populacje pod względem analizowanych cech.

Poznanie sekwencji genomu sosny zwyczajnej, czyli „planu działania” wskazującego na mechanizmy rządzące całym organizmem, pozwoli genetykom na wyselekcjonowanie genotypów o pożądanym cechach.

SUMMARY

In recent years, we have learnt more and more about the genome of another organism. In the last 10 years, there have been introduced next-generation sequencing techniques, Next Generation Sequencing (NGS), allowing the parallel sequencing reaction and generation of thousands or millions of sequences at a time.

High throughput technologies are primarily aimed at a significant reduction in the cost, but also at the speeding up of the mass sequencing process. Thanks to the development of modern sequencing techniques we are able to analyze whole genomes, with relatively small costs. Department of Silviculture, University of Life Sciences in Poznan, in cooperation with the Laboratory of Genome Analysis, A. Mickiewicz University, has an ongoing project of discovering the Scots pine genome of Kubryk population. This pine occurs in the Milicz Forest District, the area of Kubryk. In provenance studies this pine exceeds other populations in terms of the analyzed features. Understanding the pine genome sequence, the "roadmap" indicating the mechanisms governing the entire organism, will allow geneticists to select genotypes with desirable traits.